Sigma J Eng Nat Sci, Vol. 43, No. 4, pp. 1307–1320, August, 2025

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2025.00122

Research Article

Exploring heart diseases survival patterns: A multifaceted approach using survival analysis techniques on time to event data

M Darshan TEJA[1], Mokesh RAYALU G[1,*]

[1]*Department of Mathematics, School of Advanced Science, Vellore Institute of Technology, Vellore, 632014, India*

ARTICLE INFO

Article history
Received: 17 May 2024
Revised: 08 July 2024
Accepted: 05 August 2024

Keywords:
Cox-Proportional; Heart Diseases; Frailty Model; Kaplan-Meier; Survival Probability

ABSTRACT

Heart disease is a serious health issue, and effective management requires an awareness of the factors determining survival. This work used Kaplan-Meier analysis and Cox proportional hazard and frailty survival models on heart disease data. With p-values less than 0.05, our results show that factors including "Age", "Anaemia", "Creatinine Phosphokinase", "Ejection Fraction", "High Blood Pressure", "Serum Creatinine", and "Serum Sodium" are significant predictors of survival. The Cox proportional model, showed these factors effects, and their importance was highlighted by the frailty survival model using the AIC and BIC metrics. A high initial survival probability of 99.7% was revealed by Kaplan-Meier analysis, which subsequently dropped to 57.6%. With a median survival time of 111.5time units, the mean survival time was calculated to be roughly 125.19time units. These findings support risk assessment and patient treatment strategies by offering important new information on heart disease survival patterns and risk variables.

Cite this article as: Teja MD, Rayalu G M. Exploring heart diseases survival patterns: A multifaceted approach using survival analysis techniques on time to event data. Sigma J Eng Nat Sci 2025;43(4):1307–1320.

INTRODUCTION

According to the World Health Organisation, cardiovascular disease (CVD) is the primary cause of death globally, taking the lives of 17.9 million people each year. A group of conditions affecting the heart and blood vessels known as cardiovascular diseases (CVDs) include rheumatic heart disease, coronary heart disease, and cerebrovascular illness [1]. More than four out of every five fatalities from CVD are caused by heart attacks and strokes, with one-third of these deaths occurring prematurely in people under the age of 70 [2]. Tobacco use, excessive alcohol use, poor nutrition, and physical inactivity are the main behavioural risk factors for heart disease and stroke [3]. Because of their conduct, people may suffer from symptoms including HBP, HBS, high blood lipids, and obesity or overweight [4].

Frailty survival models and the Cox proportional hazards model are two complex survival analysis techniques that are combined in this inquiry to take a deep dive into the problem's severity. We may investigate how particular characteristics affect the hazard function over time by using the Cox model, which offers insightful information about their potential applications as mortality markers [5]. By

***Corresponding author.**
*E-mail address: mokesh.g@vit.ac.in

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey

adding random factors, the incorporation of frailty survival models recognizes the existence of unobserved heterogeneity among people [6]. Through the utilization of Kaplan-Meier estimates, it is possible to distinguish between the likelihood of dying from heart-related problems and from other causes [7]. This allows us to gain a more comprehensive understanding of the intricate dynamics of mortality in cases of heart illness.

This multifaceted approach aims to shed light on the multitude of patient-specific prognoses and trajectories. This approach aims to streamline the complex process of surviving cardiac disease. Through this research, we hope to gather a wealth of data that can help improve clinical practice, better classify individuals, and ultimately assist those suffering from heart issues. The primary objective of this work is to provide a comprehensive and nuanced analysis of cardiac diseases using a limited dataset that incorporates numerous parameters. These are our specific goals:

1. **Assessment of Covariate Impact:** The study uses the Cox Proportional Hazards model to determine the clinical factors that significantly impact the life expectancy of individuals with heart conditions over time.
2. **Acknowledgment of Heterogeneity:** Frailty Survival Models should be implemented to account for individual differences and hidden variables, aiming to identify and explain observed variation in survival rates among research subjects.
3. **Probabilistic Estimation:** Kaplan-Meier estimations are utilized to estimate the likelihood of dying from heart disease and other causes, aiding in a better understanding of mortality in complex settings.

LITERATURE REVIEW

Cardiovascular disease, a rapidly increasing global death rate, affects the heart and blood vessels, affecting 17.9 million people. Improve heart disease diagnosis using machine learning algorithms like SVM, KNN, NB, ANN, RF, and GDO. The GDO-based model achieved 98.54% accuracy, 99.43% sensitivity, and 97.76% precision [8]. computational intelligence techniques for heart disease diagnosis, highlighting medical feature selection as a promising technique [9]. Seven models achieve 100% accuracy using power transform, PCA, and grid search [10]. The study compares classification techniques for predicting heart disease patients, finding neural networks with the highest accuracy at 81.1%. ANN has the lowest error rate and highest accuracy [11]. Hybrid techniques [12], including vote and random forest with TPOT classifiers, improve accuracy [13]. Pooled logistic regression relates risk factors to event occurrence [14].

A study comparing Cox, Weibull, Exponential, and Frailty models in a small sample size randomised trial of Kaposi's sarcoma found the Exponential model best for fitting data, emphasising the importance of assessing fit goodness [15]. The odd Weibull family is a three-parameter Weibull distribution used for modelling survival processes, testing goodness-of-fit in data sets, particularly in censoring situations, and for simulation studies [16]. Recent cardiothoracic transplantation studies use the poly-Weibull distribution model for estimating mean survival over recipients' lifetimes, a flexible parametric model applicable to two problems using available software and inference procedures [17]. Survival analysis methods in medical research, including Kaplan-Meier estimation and log-rank tests, recommend survival regression models like Cox and AFT for continuous predictors or multiple covariates [18]. Weibull distribution, its extensions, and their applications in modelling complex lifetime data, including probability plots, parameter estimation, model validation, and future research topics [19].

A study reveals that common herd diseases like retained placenta, nonsystemic metritis, systemic metritis, ovarian cysts, and lameness decrease conception rates and increase median days open in dairy cattle [20]. Multilevel data is used in various disciplines, and three regression models are described for analysing it: Cox proportional hazards models with mixed effects, time discrete survival models, and individually exponential survival models using heart attack patient data [21]. Bayesian method to estimate Rayleigh's distribution parameters in the survival analysis of type II tuberculosis patients, revealing a value of 0.001097324 [22]. The Transformed Transformer (T-X) family generator is used to propose a new (LRD) Lomax Rayleigh distribution, which is studied for its structural properties and model fitting with simulated and real-time cancer data sets [23-25].

The significance and motivation for this research stem from the critical need to enhance predictive models for heart disease survival, considering its standing as the top cause of death worldwide. This study is inspired by the limitations of current machine learning algorithms, which frequently fail to offer precise survival probability and risk factor evaluations over time. This study uses survival analysis approaches such as Cox proportional hazard models, frailty survival models, and Kaplan-Meier analysis to provide a more precise and complete knowledge of the variables impacting heart disease outcomes. The findings, which emphasise the impact of factors like age, anaemia, and ejection fraction, offer vital insights that can help improve clinical decision-making and patient care techniques. The research thus addresses a critical vacuum in the literature by integrating strong statistical methodologies with practical cardiology applications, ultimately leading to improved patient prognosis and targeted treatment interventions.

MATERIALS AND METHODS

Cox-Proportional Hazard Model

The Cox-Proportional-Hazards Model, a statistical model used for survival analysis, was developed in 1972 by David R. Cox. This semi-parametric model assesses the risk

Table 1. An overview of cutting-edge strategies for predicting cardiovascular illnesses. Individual classifiers

Author name	Year	Data Set	Methods	Accuracy
Hassan et al. [26]	2024	Heart failure clinical records dataset (13 features)	mANN	82.95%
			FDA	82.95%
			CIT	79.55%
			BTM	78.40%
			BLR	82.95%
			NN	84.09%
			RF	85.23%
			SVM	81.82%
			RT	81.82%
			NB	79.54%
			BGLM	82.95%
			MARS	84.09%
			BoGLM	80.68%
Sara et al. [27]	2018	1. EnterprisdeData Ware house (EDW)	LR	75.4%
		2. Research Patient Data Respository(RPDR)	GB	73.9%
			Maxoutnetworks	75.4%
			DUNs	76.4%
C. Beulah et al. [28]	2019	Cleveland Heart dataset (14 features)	C4.5	79.87%
			RF	80.53%
			MP	81.52%
			BN	84.16%
			NB	84.16%
Milan et al. [29]	2011	Cleveland cardiovascular diseases dataset of 303 records (14 features)	RIPPER	81.08%
			DT	79.05%
			ANN	80.06%
			SVM	84.12%
Roohallah et al. [30]	2013	303 individual Rajaie cardiovascular (34 features)	Bagging	79.54%
			c4.5classification	68.96%

rates of different covariates under the assumption that the ratio of risks remains constant across time.

The rate of failure at time t in the given time interval is the hazard function, represented by the symbol h(t), given a probability of survival till time t. Below is a description of the Cox model:

$$h(t|Z) = h_0 t \exp(\beta_1 Z_1 + \beta_2 Z_2 + \cdots + \beta_p Z_p) \quad (1)$$

- The hazard function at time t, given the covariate values X, is $h(t|Z)$
- $h_0 t$ is the baseline hazard function, representing the hazard when all covariates are zero.
- $\beta_1, \beta_2, \ldots\ldots, \beta_p$ are the regression coefficients associated with covariates $Z_1, Z_2, \ldots\ldots, Z_p$
- $Z_1, Z_2, \ldots\ldots, Z_p$ are the covariates of predictors.
- Exp() denotes the exponential function.

The Hazard Ratio Exp(β) illustrates how the hazard changes multiplicatively with a one-unit change in the relevant covariate. The model can be used to evaluate these risk ratios and their significance.

Frailty Survival Model

Weakness to account for unobserved heterogeneity, survival models supplement the Cox Proportional-Hazards model with random effects or frailty components. When survival times are impacted by unmeasured factors, this strategy performs well. To account for individual differences, the frailty term adds a random component to the hazard function.

The Frailty survival model is expressed as follows:

$$h(t|X,Z) = \gamma_0 t \exp(\beta_1 X_1 + \beta_2 X_2 + \cdots \ldots \ldots + \beta_p X_p + \alpha Z) \quad (2)$$

- h(t|X,Z) is the HF at time t given covariate values X and a frailty term Z.

- $\gamma_0 t$ is the BHF.
- $\beta_1, \beta_2, \ldots\ldots\ldots, \beta_p$ are the regression coefficients associated with covariates $X_1, X_2, \ldots\ldots\ldots, X_3$
- α is the frailty parameter representing the random effects impact on the hazard.
- Z is the frailty term, assumed to follow a certain distribution

Kaplan-Merier

The survival function in survival analysis is estimated from observed time-to-event data using the non-parametric Kaplan-Meier estimator. It is frequently applied to material that has been suppressed and in which not every participant saw the noteworthy event.

The likelihood that a person will live past time t is represented by the Kaplan-Meier survival function, or S(t). This is how the estimator is computed:

$$\hat{S}(t) = \prod_{t_j \le t} \left(1 - \frac{d_j}{r_j}\right) \quad (3)$$

- t_j represents the observed event times.
- d_j is the No.of.events(deaths) at time t_j.
- r_j is the No.of.individuals at risk just before time t_j.

DATA SUMMARY

Data was obtained data from Kaggle dataset collection. The dataset has 299 and no missing values, therefore there is no need for imputation.

- Normal Ejection Fraction: 41%–75%; Abnormal: <41% or >75% An EF of between 50% and 75% is considered normal, according to the AHA. The range of a borderline EF is 41% to 50%.
- Serum Creatinine: 0.74 to 1.35 mg/dL (65.4 to 119.3 micromoles/L) in adult males 0.59 to 1.04 mg/dL (52.2 to 91.9 micromoles/L) for adult females
- Platelets: Between 150,000 and 450,000 platelets per microliter of blood are considered normal.
- Normal blood sodium levels range from 135 to 145 milliequivalents per litre. This is known as serum sodium.
- Creatinine Phosphkinase:10 to 120 micrograms per liter.
- Event: Patient who died. Censor: Patient is alive

RESULTS AND DISCUSSION

Cox Proportional Analysis for Heart Diseases

In the heart disease dataset, we examined the effects of several clinical factors on survival times using the Cox Proportional Hazards model.

For every covariate in the Cox Proportional Hazards model, this Table2 summarizes the following: hazard ratios, standard errors, z-values, p-values, and 95% confidence intervals. To further evaluate the overall model fit, we also offer the concordance index in addition to the outcomes of the likelihood ratio, Wald, and Score tests in Table 3.

Epidemiology results for heart diseases data by using cox proportional model

- Age is a significant predictor of heart disease risk, with each unit increase in age associated with a 4.75% higher hazard of experiencing heart disease events.
- Anaemia is a significant risk factor, with a 58.43% higher hazard ratio of heart disease events in individuals with anemia compared to non-anaemic individuals.
- Higher serum creatinine levels are linked to a 37.86% higher risk of heart disease events, according to a coefficient of 0.321 and a hazard ratio of 1.3786.
- High blood pressure significantly increases the risk of heart disease, with a 60.92% higher hazard compared

Table 2. Cox-Proportional analysis for heart diseases

Covariate	Coefficient	Hazard Ratio	Standard Error	Z Values	P value	95% CL Lower	95% CL Upper
Age	0.04641	1.0475	0.009324	4.977	6.45e-07	1.0285	1.067
Anaemia	0.4601	1.5843	0.2168	2.122	0.0338	1.0358	2.423
Creatinine Phosphokinase	0.0002207	1.0002	9.919e-05	2.225	0.026	1.00	1.00
Diabetes	0.1399	1.1501	0.2231	0.627	0.5307	0.7427	1.781
Ejection Fraction	-0.04894	0.9522	0.01048	-4.672	2.98-06	0.9329	0.972
High Blood Pressure	0.4757	1.6092	0.2162	2.201	0.0278	1.0534	2.458
Platelets	-4.635e-07	1.0000	1.126e-06	-0.412	0.6806	1.000	1.00
Serum Creatinine	0.321	1.3786	0.07017	4.575	4.76e-06	1.2014	1.582
Serum Sodium	-0.04419	0.9568	0.0237	-1.899	0.0575	0.9141	1.001
Sex	-0.2375	0.7886	0.2516	-0.944	0.3452	0.4816	1.291
Smoking	0.1289	1.1376	0.2512	0.513	0.6078	0.6953	1.861

to normal blood pressure, according to a coefficient of 0.4757 and a hazard ratio of 1.6092.

- Higher ejection fraction values are protective, decreasing the risk of heart disease events by 4.78% per unit increase, according to a negative coefficient and hazard ratio.

Age, anaemia, high blood pressure, and serum creatinine are the most significant risk factors for heart disease, with other variables like diabetes, serum sodium, sex, and smoking having less impact. Effective management through therapies and lifestyle changes can reduce risk and improve cardiovascular health outcomes.

Additionally, the Cox proportional hazards model yielded a predictive accuracy of 74%, indicating a moderately reliable performance in forecasting heart disease-related mortality (Fig. 1). Further refinement and validation of the model may enhance its accuracy and robustness in predicting outcomes.

Table 3. Metrics for different tests

	Chi squares	Degree of freedom	P Value
Likelihood ratio test	81.95	11	6e-13
Wald test	87.27	11	6e-14
Score (log rank)test	88.39	11	3e-14

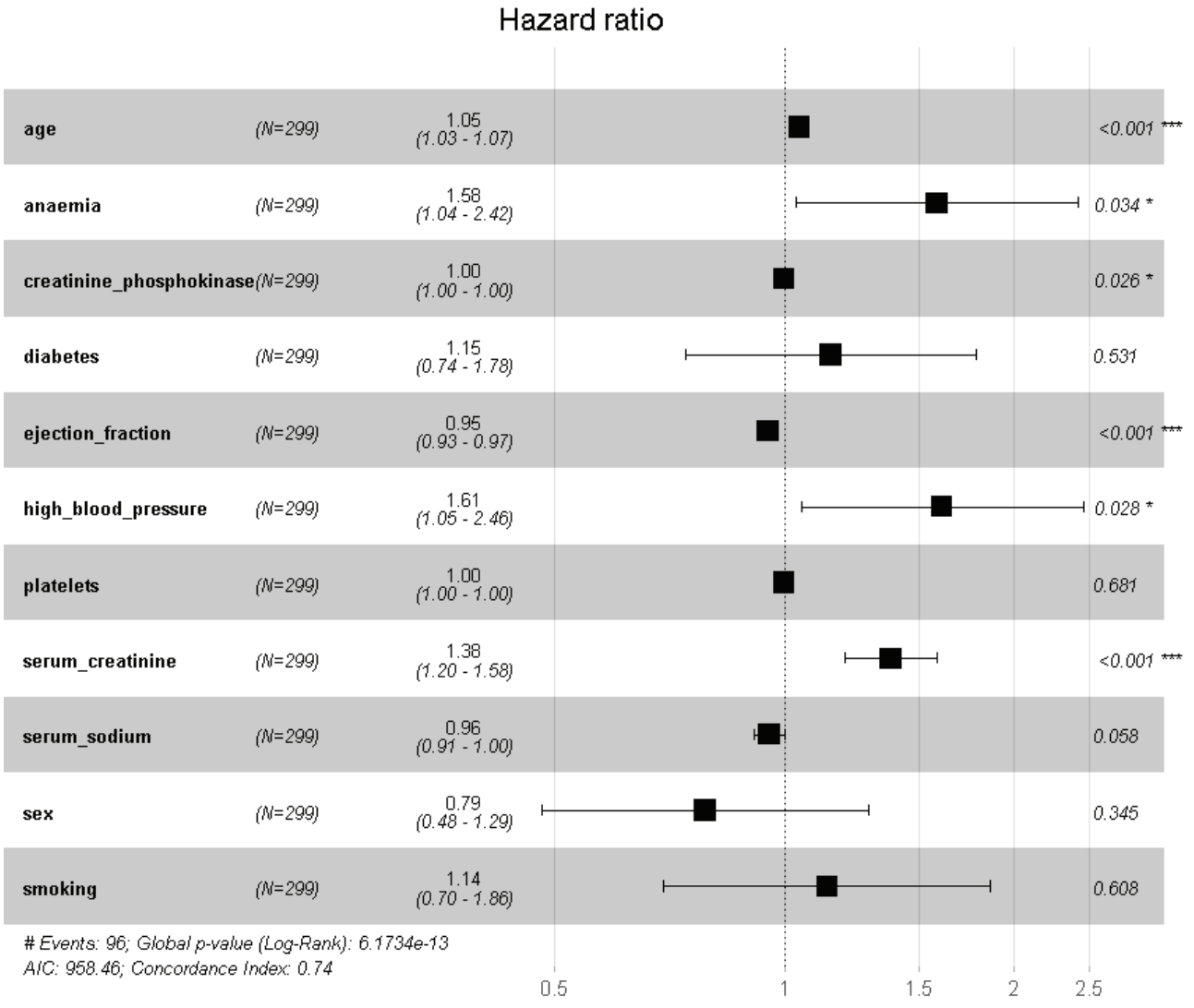

Figure 1. Cox proportional hazard graph.

Table 4. Frailty survival model analysis for heart diseases

Variable		Chi-square value	Degree of freedom	p	AIC	BIC
Age	IL	57.22	11	3.0254e-08	35.22	7.22
	PL	57.39	10	1.2122e-08	37.03	11.50
Anaemia	IL	79.37	11	1.94e-12	57.37	29.17
	PL	81.72	10	5.37e-13	60.18	32.55
Creatinine Phosphokinase	IL	78.04	11	3.5335e-12	56.04	27.83
	PL	78.11	10	1.2122e-08	58.04	32.35
Diabetes	IL	81.56	11	7.35e-13	59.56	31.35
	PL	81.57	10	2.48e-13	61.56	35.91
Ejection-fraction	IL	81.32	11	8.18e-13	59.32	31.11
	PL	110.05	10	6.439e-15	72.45	24.24
High Blood Pressure	IL	79.35	11	1.966e-12	57.35	29.15
	PL	81.74	10	5.411e-13	60.17	32.51
Serum Creatine	IL	77	11	5.6003e-12	55	26.79
	PL	133.7	10	3.44e-15	74.54	-1.30
Platelets	IL	81.78	11	6.672e-13	59.78	31.57
	PL	81.85	10	2.259e-13	61.78	36.05
Serum Sodium	IL	78.56	11	2.800e-12	56.56	28.35
	PL	78.57	10	9.623e-13	58.56	32.90
sex	IL	81.06	11	9.196e-13	59.06	30.85
	PL	81.09	10	3.112e-13	61.06	35.39
Smoking	IL	81.69	11	6.467e-13	59.69	31.48
	PL	110.05	10	2.3426e-13	61.69	36.04

From Table 3 LRT, WT, and ST (Log-Rank) collectively indicated significant differences in survival curves (Chi-Squared =81.95, df =11, p<0.001), underscoring the robustness of the analysis in capturing variations across covariates.

Frailty Survival Models

Epidemiology results for heart diseases data by using Frailty Survival model.

Based on (Table 4) chi-square values, degrees of freedom, p-values, and information criteria (AIC and BIC), we can predict the influence of various variables on heart disease and risk factor mitigation techniques.

- Age: The study's chi-square value of 57.22 indicates that age significantly impacts heart diseases, with older individuals being associated with an increased risk.
- Anaemia: A significant risk factor for heart diseases, as indicated by a chi-square value of 79.37.
- Creatinine Phosphokinase: The studies have significant chi-square values (78.04), indicating their significance in assessing heart disease risk.
- Diabetes: A significant risk factor for heart diseases, as indicated by a chi-square value of 81.56 (DF=11, p=7.35e-13).
- Ejection Fraction: The chi-square value of 81.32 emphasizes the significance of ejection fraction in evaluating heart disease risk.
- High blood Pressure:The study's findings, with a chi-square value of 79.35, highlight the significant impact of high blood pressure on heart diseases.

Table 5. SD & variances for risk factors

Variable	SD	Variances
Age	0.0199	0.003991
Anaemia	0.2822	0.0796
Creatinine Phosphokinase	0.01996	0.00039
Diabetes	8.98e-0.3	8.07e-0.5
Ejection	0.8719	0.76025
High Blood Pressure	0.2950	0.08704
Platelets	0.0199	0.0039
Serum Creatine	1.0127	1.0256
Serum Sodium	9.077e-03	8.227e-05
Sex	0.0199	0.0003
Smoking	8.9764e-03	8.057e-05

Regular health check-ups, a healthy lifestyle with a balanced diet and exercise, managing chronic conditions like diabetes and high blood pressure, and avoiding tobacco use can reduce heart disease risk factors. Early detection and management of conditions like anemia and monitoring biomarkers are also crucial.

The results of the frailty survival analysis provided insight into the differing effects of several factors on survival rates in heart disease. Notably, p-values below 0.05 indicate the statistical relevance of variables from Table 5 that significantly affect survival. The variances corresponding to these factors offer valuable insights into the degree to which they contribute to the variability in survival within the population under investigation. These results highlight the significance of taking into account a variety of parameters, such as comorbidities and physiological markers, when evaluating and forecasting survival prospects in patients with heart disease.

Figures 2-7 show frailty model graphs that show how time and important clinical parameters are related. The left side of Figure 2 shows Time in connection to Age, and the right side shows Time in relation to Anaemia. Figure 3 shows the relationship between time and creatine phosphokinase and the relationship between time and diabetes. Figure 4 shows how Time is related to Ejection Fraction and Hypertension. Figure 5 shows how Time and Platelets are related, as well as how Time and Serum Creatinine are related. Figure 6 shows how Time and Serum Sodium are related, as well as how Time and Sex are related. Figure 7 finally shows the fragility plot for Time in relation to Smoking. These charts make it easier to see how different factors affect survival time through frailty effects.

Kaplan Meier Analysis on Heart Diseases

The Kaplan-Meier survival analysis offers important information about the chances of survival for people with cardiac disease over an extended period of time from table6. At the fourth time point, the study shows a high initial survival probability of 99.7%, which progressively drops to 57.6% by the 241st time point. Each estimate's standard errors offer a measure of variability that reveals how accurate the survival probability. The range that the true survival probabilities are most likely to fall within is provided by the 95% confidence intervals. The Kaplan-Meier study, taken as a whole, shows a downward trend in survival probabilities over the observed time periods, underscoring the dynamic character of survival outcomes in heart disease patients.

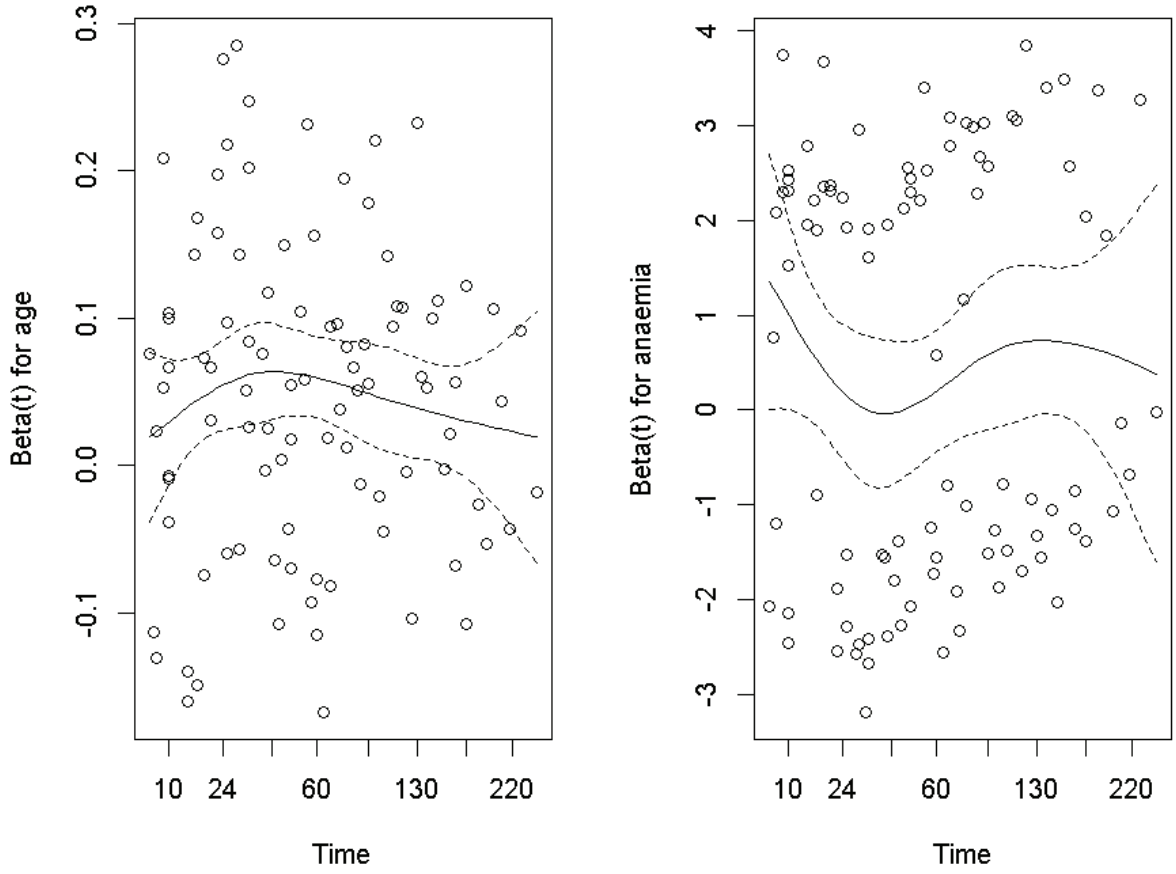

Figure 2. Frailty model plots (left) time vs age; (right) time vs anaemia.

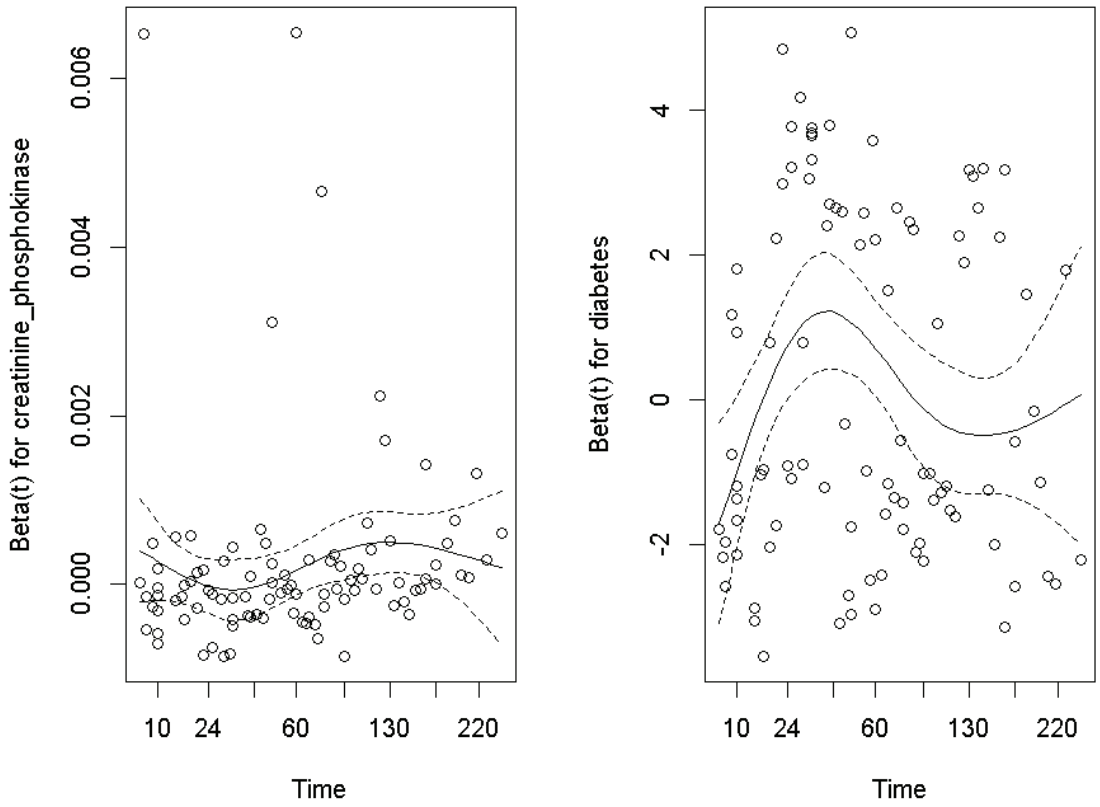

Figure 3. Frailty model plots (left) time vs creatine phosphokinase; (right) time vs diabetes.

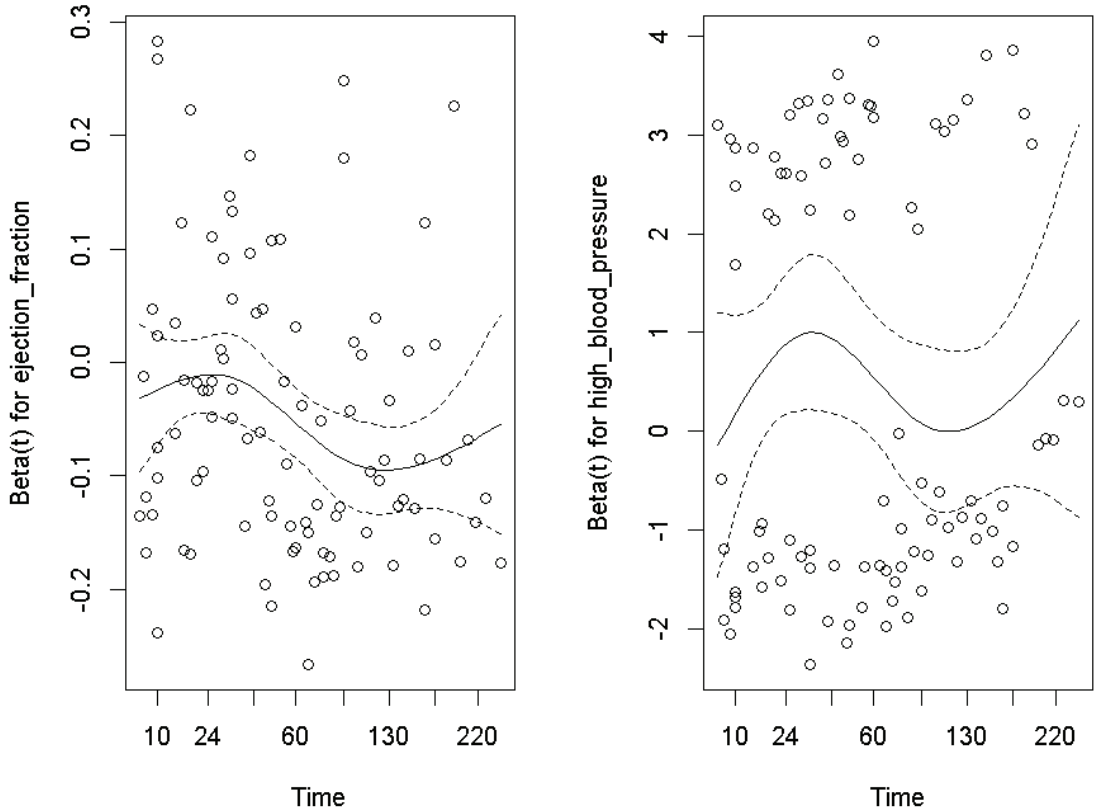

Figure 4. Frailty model plots (left) time vs ejection fraction; (right) time Vs HBP.

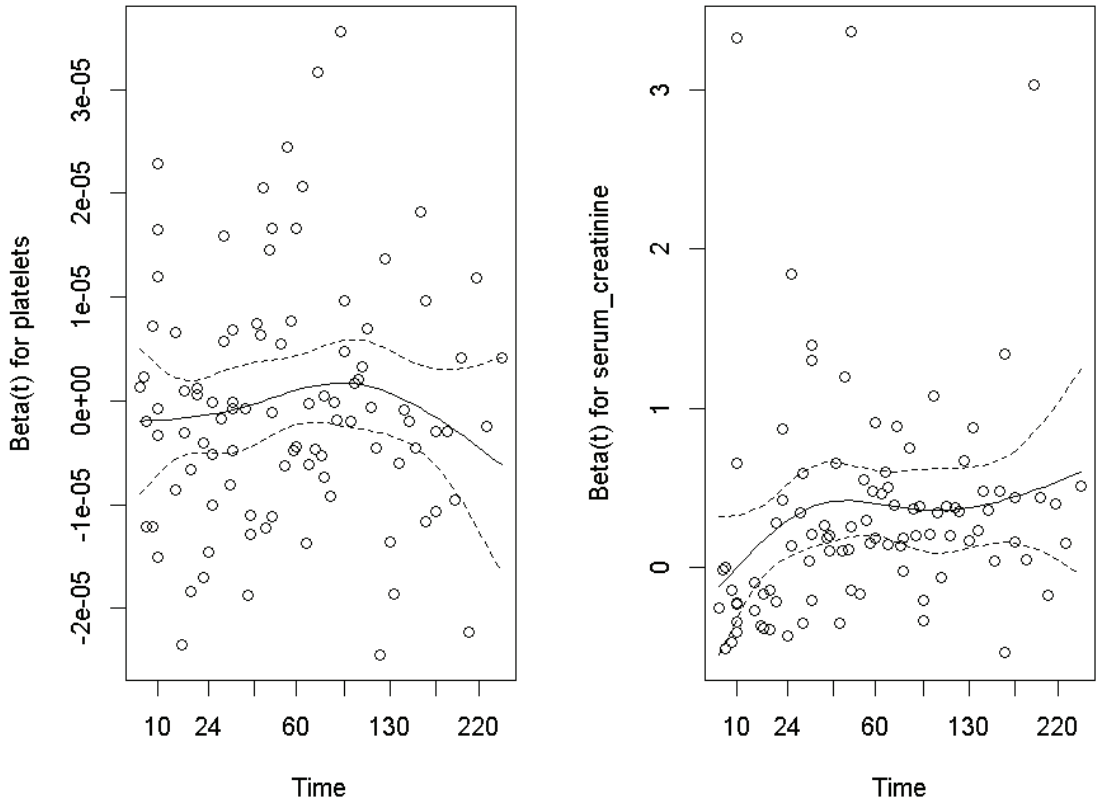

Figure 5. Frailty model plots (left) time vs platelets; (right) time vs serum creatinine.

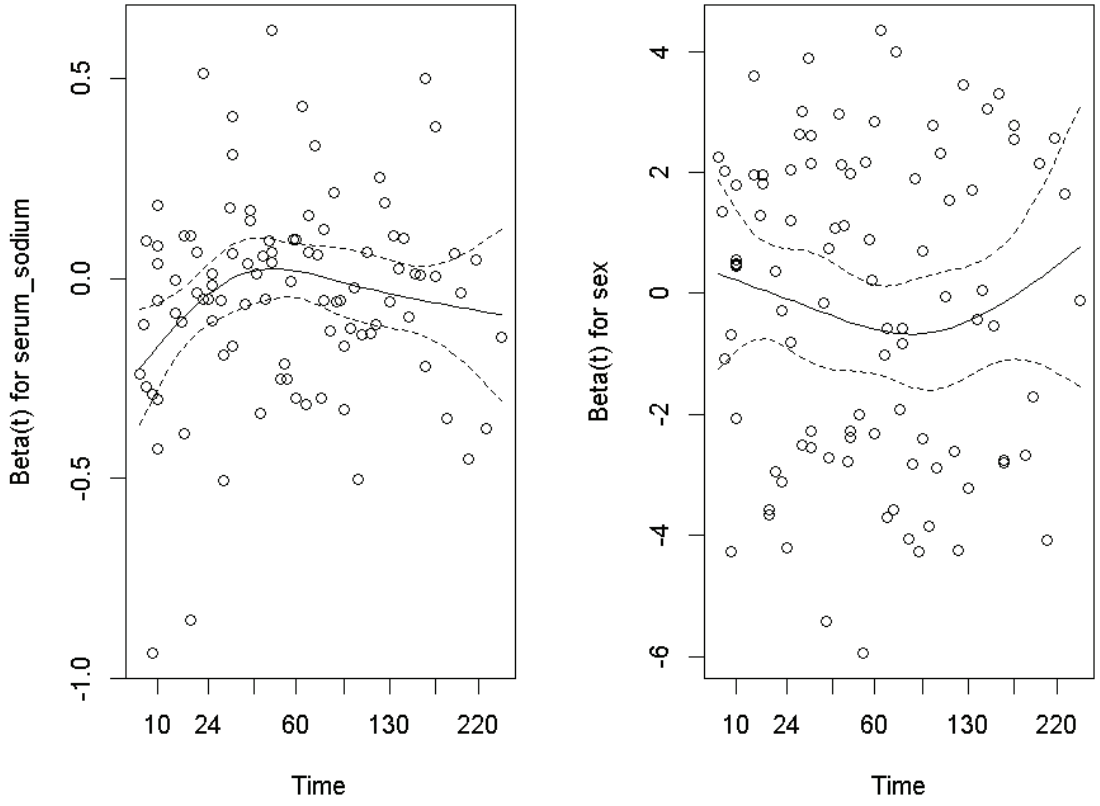

Figure 6. Frailty model plots (left) time vs serum sodium; (right) time vs sex.

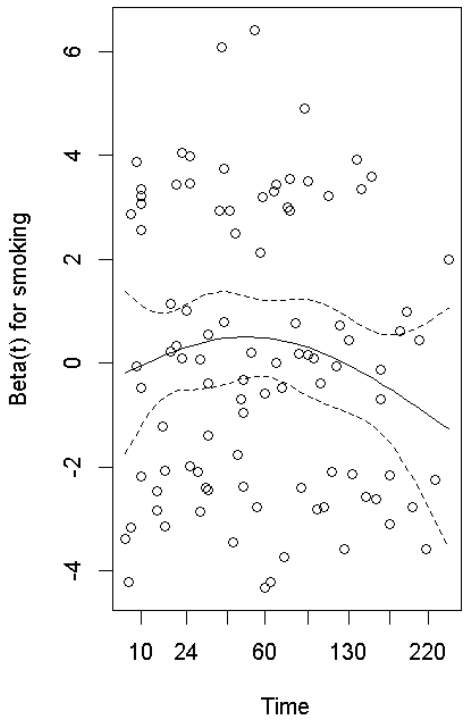

Figure 7. Frailty model plots for time vs smoking.

The Kaplan-Meier analysis yielded an estimated mean survival time of 125.19time units. Furthermore, it is computed that the median survival time is 111.5time units as shown in Figure 8. By showing the average and middle survival durations seen in the population under study, these statistics offer important insights on the central patterns of survival periods for people with heart disease.

The study examined Hassan et al. [26] findings on heart disease data using machine learning techniques like Random forest, SVM, mANN etc. The study focuses on survival approaches such as Cox proportional hazards, Kaplan-Meier estimators, and frailty models. The Cox model demonstrated that variables such as older age, anaemia, greater creatinine phosphokinase, poorer ejection fraction, high blood pressure, and elevated serum creatinine substantially increased the risk of death. Kaplan-Meier charts showed survival probability over time, but frailty models emphasized individual variations. The dataset comprised the variables "time" and "event," which are important in survival analysis. Integrating these techniques should improve predictive models and provide a better knowledge of heart disease survival.

Table 6. Kaplan Meier survival analysis

Time	No.of.Risk	No.of.Event	Survival Probability	Standard Error	Lower 95% CI	Upper 95% CI
4	299	1	0.997	0.00334	0.99	1
6	298	1	0.993	0.00471	0.984	1
7	297	2	0.987	0.00664	0.974	1
8	295	2	0.98	0.00811	0.964	0.996
10	293	6	0.96	0.01135	0.938	0.982
11	287	2	0.953	0.01222	0.93	0.977
13	284	1	0.95	0.01263	0.925	0.975
14	283	2	0.943	0.0134	0.917	0.97
15	281	2	0.936	0.01412	0.909	0.964
20	278	2	0.93	0.0148	0.901	0.959
23	275	2	0.923	0.01545	0.893	0.954
24	273	1	0.92	0.01575	0.889	0.951
26	272	3	0.909	0.01663	0.877	0.943
27	269	1	0.906	0.01691	0.873	0.94
28	268	2	0.899	0.01745	0.866	0.934
29	266	1	0.896	0.01771	0.862	0.931
30	264	4	0.882	0.01869	0.846	0.92
31	259	1	0.879	0.01893	0.843	0.917
32	258	1	0.875	0.01916	0.839	0.914
33	257	2	0.869	0.01961	0.831	0.908
35	254	1	0.865	0.01983	0.827	0.905
38	253	1	0.862	0.02004	0.823	0.902
40	252	1	0.858	0.02025	0.82	0.899

Table 6. Kaplan Meier survival analysis (*countined*)

Time	No.of.Risk	No.of.Event	Survival Probability	Standard Error	Lower 95% CI	Upper 95% CI
41	251	1	0.855	0.02046	0.816	0.896
42	250	1	0.852	0.02066	0.812	0.893
43	249	3	0.841	0.02124	0.801	0.884
44	246	1	0.838	0.02143	0.797	0.881
45	245	1	0.834	0.02161	0.793	0.878
50	244	1	0.831	0.02179	0.789	0.875
55	241	1	0.828	0.02197	0.786	0.872
59	240	1	0.824	0.02215	0.782	0.869
60	239	2	0.817	0.0225	0.774	0.863
61	236	1	0.814	0.02267	0.771	0.859
64	234	1	0.81	0.02283	0.767	0.856
65	233	2	0.803	0.02316	0.759	0.85
66	231	1	0.8	0.02332	0.755	0.847
67	230	1	0.796	0.02348	0.752	0.844
72	227	1	0.793	0.02364	0.748	0.841
73	225	2	0.786	0.02394	0.74	0.834
77	217	1	0.782	0.02411	0.736	0.831
78	216	1	0.779	0.02427	0.732	0.828
82	207	1	0.775	0.02444	0.728	0.824
88	194	1	0.771	0.02464	0.724	0.821
90	189	2	0.763	0.02504	0.715	0.813
95	180	1	0.758	0.02526	0.711	0.81
96	175	1	0.754	0.02548	0.706	0.806
100	173	1	0.75	0.02571	0.701	0.802
109	159	1	0.745	0.02597	0.696	0.798
111	155	1	0.74	0.02625	0.691	0.794
113	152	1	0.735	0.02652	0.685	0.789
115	150	1	0.73	0.02679	0.68	0.785
126	136	1	0.725	0.02713	0.674	0.78
129	135	1	0.72	0.02746	0.668	0.776
130	134	1	0.714	0.02777	0.662	0.771
135	132	1	0.709	0.02808	0.656	0.766
150	118	1	0.703	0.02848	0.649	0.761
154	117	1	0.697	0.02886	0.643	0.756
162	116	1	0.691	0.02923	0.636	0.751
170	115	1	0.685	0.02959	0.629	0.745
171	114	1	0.679	0.02993	0.623	0.74
172	113	2	0.667	0.03059	0.61	0.73
180	106	2	0.654	0.03128	0.596	0.719
193	86	1	0.647	0.03183	0.587	0.712
196	83	1	0.639	0.03238	0.578	0.706
198	79	1	0.631	0.03297	0.569	0.699
207	71	1	0.622	0.03368	0.559	0.692
214	53	1	0.61	0.03503	0.545	0.683
235	37	1	0.594	0.03776	0.524	0.673
241	33	1	0.576	0.04068	0.501	0.661

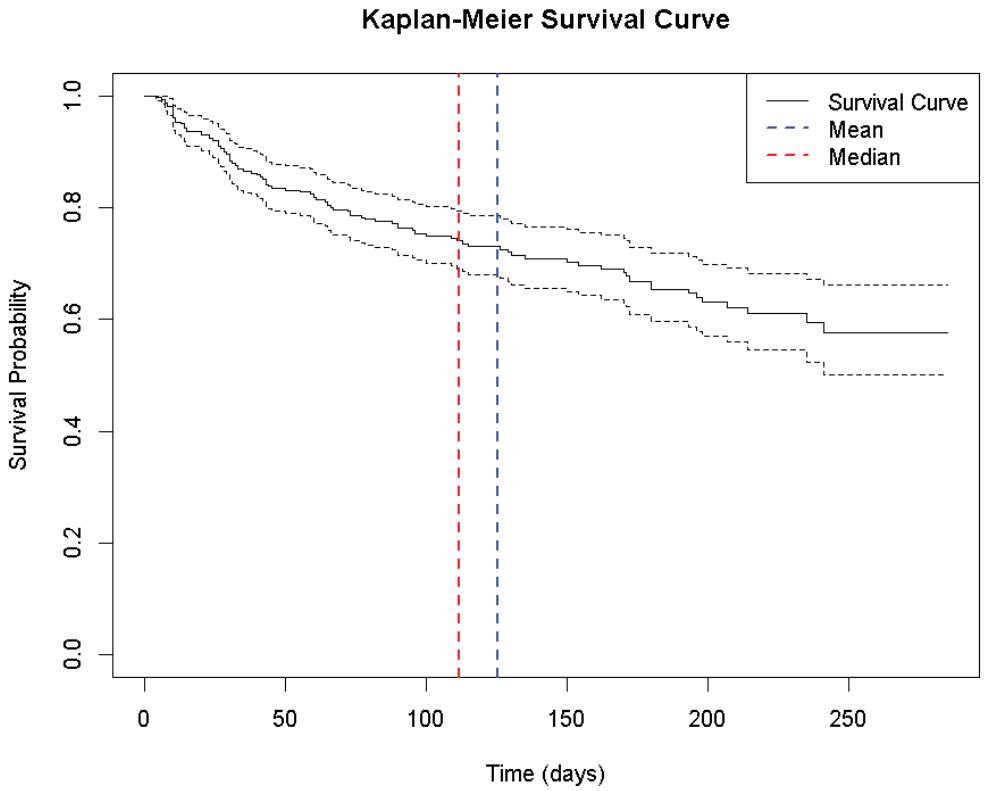

Figure 8. Kaplan-Meier Survival Curve.

CONCLUSION

Using the frailty survival model, Kaplan-Meier analysis, and Cox proportional hazard analysis performed on heart disease data, numerous important findings may be made. First, low p-values for several variables indicated that they were important predictors of survival, including age, anaemia, creatinine phosphokinase, ejection fraction, high blood pressure, serum creatinine, and serum sodium. Achieving a concordance index of 0.741, which indicates reasonably trustworthy prediction accuracy, the Cox proportional hazard model demonstrated the distinct effects of these factors on survival, along with noteworthy hazard ratios and confidence intervals.

The considerable variations in AIC and BIC measurements further supported the relevance of these factors, as indicated by the frailty survival model. Furthermore, the Kaplan-Meier analysis demonstrated how survival probabilities are dynamic and change over time, with a high survival probability initially dropping with time. The median survival time of 111.5 time units and the anticipated mean survival period of around 125.19 time units are crucial comparison points for comprehending the course of the illness. These results highlight the significance of routine health examinations, alterations in lifestyle, and efficient handling of long-term medical disorders in reducing the risk of heart disease.

To further develop knowledge and treatment of cardiac disease, future studies should investigate survival analysis distributions such as Weibull, exponential, gamma, and normal distributions, which are appropriate for accelerated failure time analysis. Enhancing patient care and predicted accuracy might be achieved by combining these methods with machine learning.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship and /or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

STATEMENT ON THE USE OF ARTIFICIAL INTELLIGENCE

Artificial intelligence was not used in the preparation of the article.

REFERENCES

[1] Samuel PO, Edo GI, Emakpor OL, Oloni GO, Ezekiel AEA, et al. Lifestyle modifications for preventing and managing cardiovascular diseases. Sport Sci Health 2024;20:23–36. [CrossRef]

[2] Ng R, Sutradhar R, Yao Z, Wodchis WP, Rosella LC. Smoking, drinking, diet and physical activity-modifiable lifestyle risk factors and their associations with age to first chronic disease. Int J Epidemiol 2020;49:113–130. [CrossRef]

[3] Manuel DG, Perez R, Sanmarth C, Taljaard M, Hennessy D, Wilson K, et al. Measuring burden of unhealthy behaviours using a multivariable predictive approach: life expectancy lost in Canada attributable to smoking, alcohol, physical inactivity, and diet. PLoS Med 2016;13:e1002082. [CrossRef]

[4] Qiu F, Li J, Gan L, Legerlotz K. Arthritis prevalence is associated with metabolic syndrome risk factors but not with physical activity in middle-aged and older patients - a cross-sectional study. BMC Geriatr 2024;24:242. [CrossRef]

[5] Lin DY, Wei LJ. The robust inference for the Cox proportional hazards model. J Am Stat Assoc 1989;84:1074–1078. [CrossRef]

[6] Shamliyan T, Talley KM, Ramakrishnan R, Kane RL. Association of frailty with survival: a systematic literature review. Ageing Res Rev 2013;12:719–736. [CrossRef]

[7] George B, Seals S, Aban I. Survival analysis and regression models. J Nucl Cardiol 2014;21:686–694. [CrossRef]

[8] Nawaz MS, Shoaib B, Ashraf MA. Intelligent cardiovascular disease prediction empowered with gradient descent optimization. Heliyon 2021;7. [CrossRef]

[9] Nahar J, Imam T, Tickle KS, Phobe Chen Y-P. Computational intelligence for heart disease diagnosis: A medical knowledge driven approach. Expert Syst Appl 2013;40:96–104. [CrossRef]

[10] Ambesange S, Vijayalaxmi A, Sridevi S, Venkateswaran. Multiple heart diseases prediction using logistic regression with ensemble and hyper parameter tuning techniques. In: Proceedings of the 2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4). IEEE; 2020. [CrossRef]

[11] Khemphila A, Boonjing V. Comparing performances of logistic regression, decision trees, and neural networks for classifying heart disease patients. In: 2010 International Conference on Computer Information Systems and Industrial Management Applications (CISIM). IEEE; 2010. [CrossRef]

[12] Amin MS, Chiam YK, Varathan KD. Identification of significant features and data mining techniques in predicting heart disease. Telem Inform 2019;36:82–93. [CrossRef]

[13] Valarmathi R, Sheela T. Heart disease prediction using hyper parameter optimization (HPO) tuning. Biomed Signal Process Control 2021;70:103033. [CrossRef]

[14] D'Agostino RB, Lee ML, Belanger AJ, Cupples LA, Anderson K, Kannel WB. Relation of pooled logistic regression to time dependent Cox regression analysis: the Framingham Heart Study. Stat Med 1990;9:1501–1515. [CrossRef]

[15] Stanley C, Molyneux E, Mukaka M. Comparison of performance of exponential, Cox proportional hazards, Weibull and frailty survival models for analysis of small sample size data. J Med Stat Inform 2016;4. [CrossRef]

[16] Cooray K. Generalization of the Weibull distribution: the odd Weibull family. Stat Model 2006;6:265–277. [CrossRef]

[17] Demiris N, Lunn D, Sharples LD. Survival extrapolation using the poly-Weibull model. Stat Methods Med Res 2015;24:287–301. [CrossRef]

[18] George B, Seals S, Aban I. Survival analysis and regression models. J Nucl Cardiol 2014;21:686–694. [CrossRef]

[19] Lai CD, Murthy DNP, Xie M. Weibull distributions. Wiley Interdiscip Rev Comput Stat 2011;3:282–287. [CrossRef]

[20] Lee LA, Ferguson JD, Galligan DT. Effect of disease on days open assessed by survival analysis. J Dairy Sci 1989;72:1020–1026. [CrossRef]

[21] Austin PC. A tutorial on multilevel survival analysis: methods, models and applications. Int Stat Rev 2017;85:185–203. [CrossRef]

[22] Elviana E, Purwadi J. Parameters estimation of Rayleigh distribution in survival analysis on type II censored data using the Bayesian method. J Phys Conf Ser 2020;1503. [CrossRef]

[23] Saritha KN, Rao GS, Rosaiah K. Survival analysis of cancer patients using a new Lomax Rayleigh distribution. J Appl Math Stat Informat 2023;19:19–45. [CrossRef]

[24] Hosmer Jr DW, Lemeshow S, May S. Applied survival analysis: regression modeling of time-to-event data. Vol 618. New York: John Wiley & Sons; 2008. [CrossRef]

[25] Ibrahim JG, Chen M-H, Sinha D. Bayesian survival analysis. Vol 2. New York: Springer; 2001. [CrossRef]

[26] Hassan MM, Zaman S, Rahman M, Bairagi AK, El-Shafai, Rathore RS, et al. Efficient prediction of coronary artery disease using machine learning algorithms with feature selection techniques. Comput Electr Eng 2024;115:109130. [CrossRef]

[27] Etu EE. The impact of machine learning algorithms on benchmarking process in healthcare service delivery. Wayne State University; 2018.

[28] Latha CBC, Jeeva SC. Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques. Informatics Med Unlocked 2019;16:100203. [CrossRef]

[29] Kumari M, Godara S. Comparative study of data mining classification methods in cardiovascular disease prediction. IJCST 2011;2:304–308.

[30] Alizadehsani R, Habibi J, Alizadeh Sani Z, Mashayekhi H, Boghrati R, Ghandeharioun A, et al. Diagnosing coronary artery disease via data mining algorithms by considering laboratory and echocardiography features. Res Cardiovasc Med 2013;2:133–139. [CrossRef]